Peer Review of Team Science Research

J. Britt Holbrook

School of Public Policy

Georgia Institute of Technology

1. Overview

This paper explores how peer review mechanisms and processes currently affect team science and how they could be designed to offer better support for team science. This immediately raises the question of how to define teams.[1] While recognizing that this question remains open, this paper addresses the issue of the peer review of team science research in terms of the peer review of interdisciplinary and transdisciplinary research. Although the paper touches on other uses of peer review, for instance, in promotion and tenure decisions and in program evaluation, the main issue addressed here is the peer review of team science research in the context of the review of grant proposals.

The paper reflects findings from prior work in this area in which the author was involved, including the National Science Foundation (NSF) sponsored Comparative Assessment of Peer Review (CAPR) project (NSF Award #0830387), which examined peer review processes at six public science and technology funding agencies around the world, and a research workshop on Transformative Research: Social and Ethical Implications (NSF Award #1129067). The paper therefore focuses on the peer review of grant proposals and the specific issues that arise when peer reviewers are asked to assess research *ex ante* – that is, before the research actually takes place. The assessment of team science during the research – *ex nunc* – and after the research has taken place – *ex post* – will also be touched on briefly.

[1] Some researchers who assert the increasing dominance of teams in knowledge production may have a minimalist definition of what constitutes a team. For instance, Wuchty, Jones, and Uzzi (2007) define a team as more than a single author on a journal article; Jones (2011) suggests that a team is constituted by more than one inventor listed on a patent. Such minimalist definitions of teams need not entail any connection with notions of interdisciplinarity, since the multiple authors or inventors that come from the same discipline would still constitute a team. Some scholars working on team collaboration suggest that factors other than differences in disciplinary background – factors such as the size of the team or physical distance between collaborators – are much more important determinants of collaboration success (Walsh and Maloney 2007). Others working on the science of teams argue that all interdisciplinary research is team research, regardless of whether all teams are interdisciplinary; research on the science of teams can therefore inform research on interdisciplinarity, as well as research on the science of team science (Fiore 2008).

The paper includes a discussion of how to deal with challenges that arise in the peer review of team science proposals, such as the challenge of assembling panels to review proposals to conduct large, multi-institution team science projects; such panels may require 20-30 reviewers to reflect the broad range of disciplinary perspectives, but it may be difficult to recruit this number of reviewers. The paper also discusses how new or revised review mechanisms could support team science (for example, NSF's CREATIV mechanism requires interdisciplinarity, and the Broader Impacts Merit Review Criterion could potentially be used to assess interdisciplinary and transdisciplinary research).

The paper reviews and synthesizes available research related to peer review processes to address the following questions:

- What are the general peer review procedures and mechanisms in federal scientific agencies, and how well-aligned are these procedures and mechanisms with the unique characteristics of team science?
- What challenges does team science pose to current peer review processes (e.g., difficulties recruiting a large enough pool of reviewers to reflect the multiple disciplines while avoiding conflicts of interest)?
- What existing peer review mechanisms (e.g., the NSF CREATIV mechanism) and/or new mechanisms (e.g., NCI funding of transdisciplinary centers) may facilitate funding or oversight of team science projects?
- What peer review mechanisms are other nations using to foster team science?
- How should peer review procedures and mechanisms be designed to facilitate the funding and effective government oversight of team science?

The paper offers options to improve the peer review of team science proposals, including conclusions on the current state of research and on further research needed to address significant gaps to improve our understanding of how peer review mechanisms and processes affect team science and how they could be changed to better support team science.

2. Introduction

Peer review is central to research. Peer review is used to make decisions about publication, about promotion and tenure, and in the evaluation of research groups, departments, programs, and universities. Peer review is also used in grant funding decisions, and almost all public research funding agencies around the world employ some form of peer review to assist in making funding decisions. When former National Science Foundation (NSF) Director Subra Suresh convened the Global Research Council in 2012, it was no accident that the topic of the first meeting was peer review.

Different agencies and research councils around the world employ peer review in different ways. Even within single agencies, there is a great deal of variation in how peer review processes are implemented. An element that has a large impact on funding decisions, regardless of the design of a peer review process, is what Cole, Cole, and Rubin (1981) termed “the luck of the reviewer draw” (p. 885). Nevertheless, agencies do design peer review processes to meet certain goals; selection of reviewers is a key element in the design process.

In fact, designing peer review processes poses several theoretical problems. Every peer review process rests on theoretical assumptions that become evident in answers to questions such as:

- Who should count as a peer?
- What criteria should be used in a given peer review process?
- How should the criteria balanced?
- What goal or goals should they serve?
- What range of interests should be represented in peer review?

Alternative answers to such questions will instantiate different values and result in the design of different peer review processes. Importantly, these designs must also be put into practice in the form of different mechanisms for peer review (Holbrook 2010a). Problems abound in moving from the theoretical design of peer review processes to the instantiation and employment of various mechanisms of peer review (Holbrook 2012). Many of these difficulties in moving from theory to practice – especially at public agencies charged both with supporting the research community and with doing so in a way that fairly and responsibly disburses public funds – involve negotiating the values of scientific autonomy and scientific accountability (Holbrook and Frodeman 2011; Holbrook and Hrotic 2013).

The most obvious question – Who should count as a peer? – has what might appear to be the most obvious answer: Academic disciplines define peers (Holbrook 2010b). However, as noted in *Facilitating Interdisciplinary Research*, scientific research “continually evolves beyond the boundaries of single disciplines” (National Research Council 2004). The most obvious answer to the question of who should count as a peer thus has an obvious flaw: Disciplinary experts are often asked to serve as peer reviewers of proposals for research that may extend beyond the bounds of their disciplinary expertise.

The idea that areas of expertise should be defined by disciplines along disciplinary lines, although widely accepted within academe (Price 1963), is itself controversial. The dominance of disciplines extends both to questions of peer review (Frodeman and Briggle 2012; Holbrook 2010b; Holbrook 2012; Holbrook and Hrotic 2013) and to the larger issue of evaluating interdisciplinary research (Huutoniemi et al. 2010; Klein 2008; Klein 2010; Wagner et al. 2011). Whether interdisciplinarity – including it variations, such as cross-, multi-, and transdisciplinarity – should be defined in terms of disciplines at all is itself open to question (Huutoniemi et al. 2010; Holbrook 2013).

The science of team science is itself an evolving field, considered by many to be a branch of science studies (Stokols et al. 2008a). Because of its potential to inform science policy, however, there are clear connections with another evolving field: the science of science and innovation policy. Unlike much of science studies, but like the science of science and innovation policy, the science of team science is pulled in the direction of answering questions posed by science policy makers rather than merely conducting research for its own sake. For this reason, much of the research in the science of team science has in fact turned toward questions of interdisciplinarity (Hall et al. 2008a) or transdisciplinarity (Stokols et al. 2008b). This paper continues this trend.

The author's research as part of the CAPR project was similarly motivated by the needs of the potential users of that research – science and technology funding agencies and policy makers. As such, CAPR represents a case study in field philosophy (Frodeman, Briggle, and Holbrook 2012). This paper also constitutes a continuation of that approach. The questions outlined in the overview, above, were asked by science policy makers. The author will do his best to answer those questions, below.

3. The Comparative Assessment of Peer Review Project: Disciplinary Autonomy and Transdisciplinary Accountability

The Comparative Assessment of Peer Review (CAPR) was a five year (2008 -2012) research project based at the University of North Texas and funded by NSF's Science of Science and Innovation Policy (SciSIP) program. CAPR examined six different funding decision procedures, including peer review of grant proposals, that incorporate societal impacts considerations across three US federal agencies: the NSF, the National Institutes of Health (NIH), and the National Oceanic and Atmospheric Administration (NOAA) and similar procedures at three non-US agencies: the European Commission (EC) Framework Programmes, focusing on the 7th (FP7), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Dutch Technology Foundation (STW).

The CAPR study had two overall objectives. First, we sought to advance our understanding of the grant proposal peer review process at public science funding bodies. Second, we hoped to improve the process of grant proposal peer review, with particular attention to identifying and helping to ameliorate difficulties that arise with the incorporation of societal impacts considerations into the peer review process.[2] As part of the CAPR project, we developed a taxonomy of different ways in which the agencies studied incorporated societal impacts considerations into the peer review of grant proposals (see Table 1).

Table 1. Comparison of 6 CAPR agencies in terms of 5 key aspects of peer review models.

Agency	Are Societal Impacts Criteria Explicit?	Do Reviewers Consider Impacts beyond Education and Outreach?	Does Agency Predetermine Expected Scope of Impacts?	Does Agency Predetermine Weighting of Impacts?	Do Reviewers Make Actual Funding Decision?
NSF	Yes	Yes	No	No	No
NOAA	Yes[a]	No[a]	Yes	Yes	No
NIH	No	Yes	No	No	No
NSERC	No	No	No	No	No
STW	Yes	Yes	Yes[b]	No	No
EC-FP7	Yes	Yes	Yes	Yes	Yes

Notes: [a] NOAA has an explicit impacts criterion that focuses only on education and outreach. Reviewers may consider societal impacts beyond education and outreach – for example, the fit with NOAA's mission or strategic plan – but not under the rubric of NOAA's education and outreach impacts criterion.

[b] STW focuses explicitly on the 'utilization' of proposed research. For a discussion of the scope of 'utilization', see the CAPR Midterm Workshop Report.

This table was originally published in Holbrook and Hrotic (2013).

All of the agencies that incorporated explicit societal impacts criteria that went beyond education and outreach activities (NSF, STW, and EC-FP7) experienced some degree of resistance to these criteria from both proposers and reviewers. In the context of the current paper, the CAPR project is relevant because these societal impacts considerations go beyond the disciplinary areas of expertise of most peer reviewers. In fact, insofar as societal impacts criteria ask about the impacts of proposed research on society, they go beyond academic disciplines altogether. Societal impacts criteria are, in essence, transdisciplinary.

One of CAPR's most fruitful comparisons was that between the EC and NSF (Holbrook and Frodeman 2011). Although both NSF and the EC placed a great deal of emphasis on impact, they designed their peer review processes quite differently. In general, NSF emphasized what could be characterized as a 'bottom-up' approach to impacts, in contrast to the EC's 'top-down' approach. NSF utilized intentionally vague criteria for "broader impacts" generically across all programs, allowing proposers a great deal of latitude to suggest their own broader impacts for peer reviewers to judge. On the other hand, the EC detailed the impacts they expected relative to specific calls for proposals. With its recent review and revision of its Merit Review Process (NSB/MR-11-22), NSF decided to continue to pursue its bottom-up approach. In fact, the revised Broader Impacts Criterion is even vaguer than the criterion that was in force from FY 1998 to FY 2012 (Holbrook 2012). This contrast between the 'bottom-up' approach using vague generic

[2] Lutz Bornmann (2013a) has a helpful recent review of the literature on societal impact assessment.

criteria to allow more room for reviewer judgment and the ‘top-down’ approach employing specific criteria for impact is also extremely relevant to the issue of designing peer review processes for the assessment of team science research, and it will be discussed further, below.

Another interesting point of comparison between NSF and the EC is the question of who is empowered to make actual funding decisions. In this respect, NSF and the EC reverse the ‘bottom-up’ vs. ‘top-down’ dynamic. Whereas the EC allows peer reviewers to make funding decisions, NSF peer reviewers are not empowered to make funding decisions. Instead, NSF peer reviewers offer their judgments on both the intellectual merit and broader impacts of proposed research; NSF program officers then take these judgments into account in making their funding recommendations to the division director, who either concurs or does not. NSF operates with what is known as a ‘strong program officer’ modeled largely on the US Office of Naval Research. The design of peer review processes for the *ex ante* assessment of team science research should also take the issue of who actually makes funding decisions into account.

Despite these differences in the design of their peer review models and their funding decision procedures, both NSF and the EC encountered resistance from proposers and peer reviewers to transdisciplinary societal impacts criteria. This resistance was most often expressed by the suggestion that societal impacts criteria are ‘unclear’. This raised the possibility that, insofar as such transdisciplinary criteria go beyond the realm of proposers’ and peer reviewers’ areas of disciplinary expertise, both felt unqualified to speak to the broader societal impacts of research.

The EC responded to such suggestions by specifying impacts, making them explicitly relevant to particular calls (Holbrook and Frodeman 2011). NSF also considered moving in the same direction; a list of ‘National needs’ was put forward in a draft revision of NSF’s Merit Review process (Holbrook 2012). However, research conducted as part of the CAPR project raised the possibility that proposers and reviewers in fact felt qualified to speak to questions of broader societal impacts (Holbrook and Hrotic 2013). This suggested claims that societal impact criteria were ‘unclear’ should not necessarily be taken at face value. Instead, such complaints might be expressions of unwillingness to consider such transdisciplinary criteria as important factors in funding decisions. Put differently, complaints from scientists about societal impacts criteria might be expressions of the demand for scientific autonomy in the face of growing demands for scientists to demonstrate accountability to society (Frodeman and Holbrook 2011).

The discussion in this section of the paper suggests a tension between disciplinary autonomy and transdisciplinary accountability. Of course, things are not so simple. It is possible to think of disciplinary autonomy itself in terms of accountability – disciplinary autonomy could be analyzed in terms of accountability to one’s disciplinary peers. Considering disciplinary autonomy in terms of accountability to one’s disciplinary peers raises many issues for the design of processes for the peer review of team science research. These issues are discussed §4, below.

4. Interdisciplinary Peer Review: Respect for Disciplinary Expertise and Autonomy

Disciplines define peers, and peer review is often designed to uphold disciplinary standards – of rigor, of method, of subject matter, and generally of what counts as good research within a discipline. When a piece of research is subject to peer review, then, it typically means that disciplinary standards will determine whether it passes muster to be published (in the case of a manuscript submitted for publication) or to be funded (in the case of a grant proposal). If peer review depends on disciplinary standards, then how is it possible to review proposals that go beyond disciplinary bounds?

Decisions regarding promotion and tenure typically involve a larger body of work; but this is also typically subject first and foremost to disciplinary peer review (by peers within the department and external referees, who are typically scholars of high standing within the discipline). Tenure decisions usually also involve review by members of the faculty from disciplines other than that of the person up for tenure review. These tenure review committees tend to rely heavily on the reports of the disciplinary peers within the department and the external disciplinary reviews. The largest factor in their decisions, however, remains the candidate's record of *peer reviewed publications* (National Research Council 2012, Harley 2013). Such publications ideally appear in the top journals within the researcher's field of expertise. In other words, non-disciplinary 'peers' place their trust in the judgment of disciplinary peers.

This sort of respect shown by members of review panels for the disciplinary expertise of other reviewers is also sometimes evident in the peer review of grant proposals. Lamont, Mallard, and Guetzkow (2012) identify several 'rules' adopted by panelists, including "deferring to expertise" and "respecting interdisciplinary sovereignty" (p. 431). Lamont (2009) provides a way of viewing the process of peer review – as an interactive social process in which the participants aim at a kind of Habermasian ideal speech situation, in which reviewers from different disciplines respect each other's differing disciplinary standards and aim to reach a consensus decision – that may prove useful in the review of interdisciplinary grant proposals. She also suggests that more intensive training of personnel at public funding institutions may be necessary in order to sensitize agencies to the exigencies of evaluating interdisciplinary research.

Research that suggests respect for disciplinary diversity may not address other biases, such as sexism or nepotism (Wennerås and Wold 1997). However, it does suggest that the most typical practice in terms of the peer review of grant proposals – single blinding, in which reviewers know and consider the identity and qualifications of the proposer – may not affect disciplinary diversity. In prepublication peer review of journal manuscripts – typically double-blind, so that neither the reviewer nor the author know the other's identity – disciplinary identities can be revealed in terms of methods employed (or lack thereof), literature cited, and so forth. Moreover, reviewers of journal manuscripts are often asked to comment on whether a manuscript is a good fit for the journal – a question that is often justifiably determined on disciplinary or even subdisciplinary grounds.

5. Disciplining Interdisciplinarity or Dedisciplining Disciplines?

The question of rigor looms large over interdisciplinary and transdisciplinary research. Those who engage in interdisciplinary or transdisciplinary research have responded to the question of rigor in various ways. One of the most interesting – and controversial – is to suggest that researchers *on* interdisciplinarity should *discipline* themselves. Taken to an extreme, some researchers on interdisciplinarity have even suggested the need to form a discipline of interdisciplinarity; Gabriele Bammer's argument for Integration and Implementation Sciences (2013) stands out in this regard. In just the way that other disciplines produce peers, Bammer's approach would generate *experts* in interdisciplinarity and transdisciplinarity.

Following Bammer's suggestion would have interesting implications for both the review of team science research proposals and the science of team science in general. The latter question – the degree to which the science of team science should become a discipline – is beyond the scope of this paper. Nevertheless, it is a pressing issue for future research and discussion. But, using the premise of a discipline of interdisciplinarity, it is possible to conduct a thought experiment regarding the peer review of team science research.

Suppose we had a discipline of interdisciplinarity, that is, full-fledged *experts* in the field. It would then make sense to draw from this reservoir of expert knowledge in order to select reviewers for team science research proposals. But insofar as these expert interdisciplinarians are experts in interdisciplinarity itself, rather than in some (other) disciplinary field of research, we are faced once again with the question of whether these experts are in any sense *peers* of the proposers of the team science research. To the expert interdisciplinarians, experts in biology, geology, and physics proposing to engage in team research about the origins of life on earth might appear rank amateurs when it comes to interdisciplinarity. Even retreating into disciplinary standards of rigor – supposing astrobiology were a full-fledged discipline – would not solve this problem. Indeed, the astrobiologists would be well-justified in arguing that non-astrobiologists were not their peers and hence had no business judging astrobiology proposals. Were we to fall back on norms such as 'deferring to expertise', we could imagine teams that included expert interdisciplinarians as well as expert astrobiologists; review panels could have a similar mix of expertise.

At the other extreme – also interesting and controversial – is the idea that engaging in interdisciplinary and transdisciplinary research requires disciplinary experts to *dediscipline* themselves. Robert Frodeman's (2013) *Sustainable Knowledge: A Theory of Interdisciplinarity* presents a striking argument for taking a dedisciplined approach, suggesting that the 'age of disciplinarity' is coming to an end. Another thinker who shares Frodeman's relative position to the disciplinary interdisciplinarians is Steve Fuller, who argues in favor of what he terms 'deviant' interdisciplinarity (2010). The deviant interdisciplinarian is one who explores the paths not taken by normal interdisciplinary research, attempting to broaden the scope of what counts as an interesting question.

Fuller and Frodeman have their differences – most obviously in terms of the value of the pursuit of more knowledge, where Fuller favors *more* and Frodeman *less*. But they are united in both their distance from the disciplinary interdisciplarians and in viewing peer review itself as a force that perpetuates the disciplinary pursuit of knowledge (Frodeman and Briggle 2012, Fuller 2012). For both Fuller and Frodeman, the question 'Who should count as a peer?' suggests more than simply carefully selecting reviewers so that the disciplines in the team science proposals were reflected in the makeup of the reviewers; redesigning peer review processes and mechanisms themselves might be in order.

6. Mechanisms Designed for the Review of Interdisciplinary and Transdisciplinary Proposals

As discussed above, different designs for peer review processes exist at different funding agencies, and there are differences even within agencies, sometimes because of different mechanisms and sometimes because of the way that these mechanisms are implemented. Even within a single program in a single agency soliciting proposals aimed at a single discipline, changing the specific reviewers is likely to affect the process of peer review. As with any process of design, it is always possible for users of the mechanism to employ it in a way the designers would classify as a *mis*use. For this reason, any classification of different peer review processes into categories of what those mechanisms were 'designed for' cannot serve to predict accurately how those mechanisms will actually be used.

Put differently, there are very few 'laws' of peer review that, if followed, will result in a flawless process.

Designing mechanisms for peer review thus takes on an experimental character in which the best course of action is often to learn by doing, taking limited risks, and modifying peer review processes based on what was learned. Learning takes place among proposers and reviewers, as well as among agency staff. Unconsidered, wholesale changes to a peer review process by those lacking experience with the process, although often possible, are rarely wise – especially when judged in terms of efficiency and effectiveness.

Here, I briefly discuss several different approaches designed to deal with interdisciplinary or transdisciplinary research. One point to note immediately is that peer review mechanisms designed for interdisciplinary research often need to encourage more risk-taking on the part of proposers, reviewers, and agency staff (Langfeldt 2006). This consideration is evident in NSF's CREATIV mechanism, which was designed to fund 'potentially transformative' – or 'high risk, high reward' – research (Frodeman and Holbrook 2012).[3] One interesting choice NSF made was to shift the review of CREATIV proposals from external peers to internal merit review (conducted by NSF staff only) in an effort to avoid the conservatism of the peer community. Another is to require review by at least two intellectually distinct programs, i.e., a sort of interdisciplinarity. Both design considerations fall on the side of the 'dedisciplinary' extreme outlined in §5, above.[4]

NSF's CREATIV mechanism is not the only process designed to support high risk research, even within NSF. After attempting an environmental scan of US federal agency support for such research, however, the President's Council of Advisors on Science and Technology (PCAST) concluded that current efforts represented merely a few drops in the bucket:

> While this plethora of initiatives, each worthy in its own way, gives an illusion of significant progress [in funding revolutionary research], in truth the sum of all of these programs is tiny, almost invisible, in comparison to each agency's dominant model [of funding evolutionary research]. (PCAST 2012, p. 70)

Because of this conclusion that not enough has been done to support transformative research, PCAST recommends the following:

> In addition to specific programs focused on supporting new and emerging areas of research, agencies have developed review criteria and other policies to target funding for ground-breaking, high-reward projects. In our estimation, however, none of these has been sufficient to the magnitude of the problem. We call for a substantially larger effort to support research proposals (1) with potential game-changing impact; (2) that fall outside traditional disciplines; and (3) that are people, rather than project, based. (PCAST 2012, p. 71)

This third recommendation is even further along the spectrum toward the 'dedisciplinary' extreme.

---

[3] The contributions of Evans and Rogers contained in the report are especially salient.

[4] See Luukkonen (2012) for a discussion of the European Research Council's approach to risk and reward in funding excellent research.

An effort that falls more toward the 'disciplinary' extreme is the NIH National Cancer Institute's Transdisciplinary Research on Energetics and Cancer (TREC) initiative. Hall et al. (2008b) outline a protocol (cancercontrol.cancer.gov/trec/TREC-Protocol-

2006-09-27.pdf) developed for "evaluating the integrative qualities and scope" of pilot project grant proposals during the first year of the TREC initiative. Here, the focus was less on making sure that peer review did not quash risk-taking and more on making sure that the criteria developed as part of this funding mechanism were suited to the review of interdisciplinary and transdisciplinary research.

The protocol (based on Mitrany and Stokols 2005) included criteria for:

- *Disciplines represented* in the proposal
- *Type* of cross-disciplinary integration (uni-, multi-, inter-, or transdisciplinary) based on Rosenfield (1992)
- *Scope* of transdisciplinary integration (rated on a 10-point Likert scale ranging from 'none' to 'substantial')
- The overall scope of the proposal (rating other criteria together for an overall score, also on a 10-point Likert scale ranging from 'none' to 'substantial')

These criteria are justified with reference to *the literature* on team science and interdisciplinarity (that is, the same way disciplinary methods and protocols are justified).

Another interesting feature of the TREC pilot proposal review mechanism is the actual process according to which reviews were conducted. Hall et al. (2008b) describe the process as follows:

> The reviewers were trained by members of the evaluation team to ensure consistent interpretations and applications of the written-products rating scales. Consensus conference calls were later held with a moderator and members of the NCI evaluation team. Members of the evaluation team included individuals with a wide range of cross-disciplinary clinical and research experience, as well as previous experience conducting evaluations of other large transdisciplinary initiatives. Discrepant scores on the various rating scales for each proposal were discussed among the group until consensus was reached. (p. S168)

After the review process took place, inter-rater reliabilities were assessed. In other words, something resembling disciplinary rigor was established, maintained, and assessed for the review of interdisciplinary grant proposals.

Establishing a rigorous peer review protocol in designing the mechanism for reviewing interdisciplinary and transdisciplinary proposals is a labor intensive process. At larger scales, it may be prohibitively difficult to accomplish. As Hackett and Chubin (1990) note, different processes of peer review often represent tradeoffs in terms of different values – in this case, rigor and efficiency. Are there sufficient experts in team science, and will enough be willing to serve as reviewers under such a strict regime, in order to make such a rigorous mechanism viable?

There are also various ways of dealing with transdisciplinary proposals. In addition to NSF's Broader Impacts Criterion, STW (also mentioned in §3, above) incorporates a 'users' committee as part of its review process, which it refers to as an 'ongoing conversation' (see Holbrook 2010a for a fuller description). The US Congressionally Directed Medical Research Program (CDMRP) also incorporates users – patients or family members of patients – into its review process. All of these mechanisms are located on the 'dedisciplinary' side of the spectrum to varying degrees.[5]

7. Policy Options

Tinkering with peer review mechanisms is inherently dangerous. Once proposers, reviewers, and funding agency officials learn a process, there is always resistance to change. Often, this resistance is well-justified. However, if funding agencies want to fund more interdisciplinary and transdisciplinary research, then they will also have to brave this danger.

Since tinkering with peer review mechanisms is inherently dangerous, science funding agencies should consider carefully what goal they want to achieve, why they want to achieve it, how they plan to do so, how they will know if they succeed, and what benefits could accrue if they are successful. Agencies should also decide how much they are willing to risk on interdisciplinary and transdisciplinary proposals. Risks of not acting should also be considered.

One way to mitigate risks is to look carefully at what other science funding agencies have done.[6] Another way to mitigate risks is to consult the research on interdisciplinarity and transdisciplinarity. Although no identifiable discipline of interdisciplinarity currently exists, whether we should think about forming one is a question that is increasingly being addressed in

[5] This tension between disciplinary standardization and calls for dedisciplined, decentralized approaches is also currently playing itself out now in two developing areas of *ex nunc* and *ex post* evaluation: bibliometrics and altmetrics. Although bibliometrics have been around for a while, new techniques are being developed, and there is increasing pressure to use bibliometrics for the assessment of individuals. Altmetrics is a fledgling field that develops tools for monitoring social media attention to scholarly products. Both bibliometricians concerned about individual level metrics and altmetricians are currently in serious talks about developing 'standards'. Lutz Bornmann (2013) suggests that there is currently a scientific revolution going on in scientometrics.

[6] In my opinion, the recent Review and Revisions of NSF's Merit Review Process (NSB/MR-11-22) is an outstanding example of how things ought to be done on a large, foundation-wide scale.

the literature on interdisciplinarity. And, even if there is as yet no discipline, there *is* a literature – though the precise borders of it, and the areas of overlap and contestation, remain unclear. Research on the evaluation of interdisciplinary and transdisciplinary research, *ex ante*, *ex nunc*, and *ex post*, exists; much of it is cited here.

Another way to mitigate risks would be to engage with researchers on questions policy makers want to see answered. Not all researchers will be interested in answering such questions, of course. Disciplinarians have a tendency to want to set their own research agendas, as well as a tendency to end up talking only among themselves. Dedisciplinarians, although often more open to engagement with non-academics on their own terms, are also in some ways riskier prospects. Insofar as they have an ambivalent attitude toward the questions – and the rigor – of their day, their reputations among their (disciplinary) peers may suffer.

8. Areas for Further Research

Given the local variability of peer review processes even within funding agencies, the search for a particular *method* for the ex ante peer review of interdisciplinary and transdisciplinary research may be in vain. On the other hand, even if the search for inter-rater reliability in peer review of grant proposals appears forced, it cannot simply be a matter of 'all is permitted'. Is there a range, or even a sweet spot, between the Wild West and rigor mortis? What is the proper balance between autonomy (accountability to oneself or one's peers) and accountability to society? How could we measure that? What do peers value when they rate the same proposals – even *disciplinary* proposals – differently? Can we leverage – rather than trying to eliminate – bias to ensure that interdisciplinary and transdisciplinary proposals review well? Should we do so? What is the relation between encouraging interdisciplinary/transdisciplinary/transformative/team science at the funding agency level and promotion and tenure criteria at the department, college, or university levels? Are there alternatives to traditional peer review that ought to be considered for interdisciplinary and transdisciplinary proposals? Should these decisions be made on the basis of theory, 'local' practice, experiment, or some combination thereof? If we treat variations on existing peer review processes as experiments, how long should we run them? How can we connect *ex ante*, *ex nunc*, and *ex post* review of team science research in ways that enrich evaluation overall? Is research evaluation worth the cost?

9. Conclusion

Although I admittedly tend toward the dedisciplined extreme myself, there do have to be some standards according to which we operate (Holbrook, Barr, Brown 2013). The most interesting question is whether we must know what those standards are *before* we act. But this is a philosophical question, perhaps – that is, it is a question the answer to which is not a matter of expertise and which no further research can answer. In any case, one thing seems to me certain:

The questions we find interesting and the answers we give say something about who we are as people. At least they reveal what it is we value. The same can be said of funding agencies.

References Cited

Bammer, Gabriele (2013). *Disciplining Interdisciplinarity: Integration and Implementation Sciences for Researching Complex Real-World Problems*. ANU E Press. Available at <http://epress.anu.edu.au?p=222171>, last accessed 21 October, 2013.

Bornmann, L. (2013), What is societal impact of research and how can it be assessed? a literature survey. J. Am. Soc. Inf. Sci., 64: 217–233. doi: 10.1002/asi.22803.

Bornmann, Lutz (2013b). Is there currently a scientific revolution in scientometrics?. arXiv preprint arXiv:1307.6307.

CAPR Midterm Workshop Report. Available at <http://csid-capr.unt.edu/fedora/repository/capr:675/OBJ/>, last accessed 22 September 2013.

Cole, Stephen, Jonathan R. Cole, and Gary A. Simon (1981). "Chance and consensus in peer review." *Science* 214, no. 4523: 881-886.

Chubin, Daryl E., and Edward J. Hackett (1990). *Peerless science: Peer review and US science policy*. SUNY Press.

Fiore, Stephen M. (2008). "Interdisciplinarity as teamwork how the science of teams can inform team science." *Small Group Research* 39.3: 251-277.

Frodeman, Robert, and Adam Briggle. "The dedisciplining of peer review." Minerva 50, no. 1 (2012): 3-19.

Frodeman, Robert, Adam Briggle, and J. Britt Holbrook (2012). "Philosophy in the Age of Neoliberalism." *Social Epistemology: A Journal of Knowledge, Culture, and Policy*. 26 (2-4): 1-20. DOI:10.1080/02691728.2012.722701.

Frodeman, Robert and J. Britt Holbrook (2012). *The Promise and Perils of Transformative Research*. UNT Digital Library: http://digital.library.unt.edu/ark:/67531/metadc84363/.

Robert Frodeman and J. Britt Holbrook (2011). "NSF's Struggle to Articulate Relevance." *Science*, vol. 333, July 8: 157-58.

Fuller, Steve (2010). "Deviant Interdisciplinarity," in *The Oxford Handbook of Interdisciplinarity*, Robert Frodeman, Julie Thompson Klein, Carl Mitcham, eds. Oxford: Oxford University Press: 50-64.

Fuller, Steve (2012). "Peer Review: A Theory in Search of Best Practice?" in *Peer Review, Research Integrity, and the Governance of Science – Practice, Theory, and Current Discussions*.

Robert Frodeman, J. Britt Holbrook, Carl Mitcham, and Hong Xiaonan. Beijing: People's Publishing House: 521-540.

Hall, Kara L., Annie X. Feng, Richard P. Moser, Daniel Stokols, and Brandie K. Taylor (2008a). "Moving the science of team science forward: collaboration and creativity." *American journal of preventive medicine* 35, no. 2: S243-S249.

Hall, Kara L., et al. (2008b). "The collaboration readiness of transdisciplinary research teams and centers: findings from the National Cancer Institute's TREC Year-One evaluation study." American journal of preventive medicine 35.2: S161-S172.

Harley, Diane (2013). "Scholarly Communication: Cultural Contexts, Evolving Models." *Science* 4 October: 80-82.

Holbrook, J. Britt (2010a). "The use of societal impacts considerations in grant proposal peer review: A comparison of five models," in *Technology and Innovation*, Vol. 12: 213–224. DOI: 10.3727/194982410X12895770314078.

Holbrook, J. Britt (2010b). "Peer Review," in *The Oxford Handbook of Interdisciplinarity*, Robert Frodeman, Julie Thompson Klein, Carl Mitcham, eds. Oxford: Oxford University Press: 321-32.

Holbrook, J. Britt (2012). "Re-assessing the science – society relation: The case of the US National Science Foundation's broader impacts merit review criterion (1997 – 2011)," in *Peer Review, Research Integrity, and the Governance of Science – Practice, Theory, and Current Discussions*. Robert Frodeman, J. Britt Holbrook, Carl Mitcham, and Hong Xiaonan. Beijing: People's Publishing House: 328 – 62.

Holbrook, J. Britt (2013). "What is interdisciplinary communication? Reflections on the very idea of disciplinary integration." *Synthese*, 190 (11): 1865-1879. DOI: 10.1007/s11229-012-0179-7.

Holbrook, J. Britt and Robert Frodeman (2011). "Peer review and the *ex ante* assessment of societal impacts." *Research Evaluation*, 20 (3): 239–246. DOI: 10.3152/095820211X12941371876788.

Holbrook, J. Britt, Kelli R. Barr, and Keith Wayne Brown (2013). "Research Impact: We Need Negative Metrics Too." Nature 497, 439 (23 May 2013) doi:10.1038/497439a Published online 22 May.

Holbrook, J. Britt and Steven Hrotic (2013). "Blue skies, impacts, and peer review." *ROARS Transactions: A Journal of Research, Policy, and Evaluation* 1 (1). Published online 21 July. DOI: 10.13130/2282-5398/2914.

Huutoniemi, Katri (2010). "Evaluating interdisciplinary research," in *The Oxford Handbook of Interdisciplinarity*, Robert Frodeman, Julie Thompson Klein, Carl Mitcham, eds. Oxford: Oxford University Press: 309-320.

Huutoniemi, Katri, Julie Thompson Klein, Henrik Bruun, and Janne Hukkinen (2010). "Analyzing interdisciplinarity: Typology and indicators." *Research Policy* 39, no. 1: 79-88.

Jones, Benjamin F. (2011). "As science evolves, how can science policy?" *Innovation Policy and the Economy*, Volume 11. University of Chicago Press. 103-131.

Klein, Julie T. (2008). "Evaluation of interdisciplinary and transdisciplinary research: a literature review." *American journal of preventive medicine*, 35(2), S116-S123.

Klein, Julie T. (2010). "A taxonomy of interdisciplinarity," in *The Oxford Handbook of Interdisciplinarity*, Robert Frodeman, Julie Thompson Klein, Carl Mitcham, eds. Oxford: Oxford University Press: 15-30.

Lamont, Michèle (2009). *How professors think: inside the curious world of academic judgment*. Cambridge, MA: Harvard University Press.

Lamont, Michèle, Grégoire Mallard, and Joshua Guetzkow (2012). "Beyond Blind Faith: Overcoming Obstacles to Interdisciplinary Evaluation," in *Peer Review, Research Integrity, and the Governance of Science – Practice, Theory, and Current Discussions*. Robert Frodeman, J. Britt Holbrook, Carl Mitcham, and Hong Xiaonan. Beijing: People's Publishing House: 415-453.

Langfeldt, Liv (2006). "The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments." *Research Evaluation*, volume 15, number 1, April 2006, p.p. 31–41.

Luukkonen, T. (2012). Conservatism and risk-taking in peer review: Emerging ERC practices. Research Evaluation 21(1): 48-60. DOI: 10.1093/reseval/rvs001.

Mitrany M, Stokols D. (2005). Gauging the transdisciplinary qualities and outcomes of doctoral training programs. J Plan Educ Res; 24: 437– 49.

National Research Council (2004). *Facilitating Interdisciplinary Research*. Washington, DC: The National Academies Press.

National Research Council (2012). *For Attribution – Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop*. Washington, DC: The National Academies Press.

NSB/MR-11-22. National Science Foundation's Merit Review Criteria: Review and Revisions. https://csid-capr.unt.edu/fedora/repository/capr:1176.

PCAST 2012. Transformation and Opportunity: The Future of the US Research Enterprise 2012. Available online at: http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast_future_research_enterprise_20121130.pdf. Last accessed 31 January, 2013.

Price, Derek. J. de Solla (1963). *Little science, Big science*. New York: Columbia University Press.

Rosenfield PL (1992). The potential of transdisciplinary research for sustaining and extending linkages between the health and social sciences. Soc Sci Med; 35: 1343–57.

Stokols, Daniel, Kara L. Hall, Brandie K. Taylor, and Richard P. Moser (2008a). "The science of team science: overview of the field and introduction to the supplement." *American journal of preventive medicine* 35, no. 2: S77-S89.

Stokols, Daniel, Shalini Misra, Richard P. Moser, Kara L. Hall, and Brandie K. Taylor (2008b). "The ecology of team science: understanding contextual influences on transdisciplinary collaboration." *American journal of preventive medicine* 35, no. 2: S96-S115.

Wagner, Caroline S., J. David Roessner, Kamau Bobb, Julie Thompson Klein, Kevin W. Boyack, Joann Keyton, Ismael Rafols, and Katy Börner (2011). "Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature." *Journal of Informetrics* 5, no. 1: 14-26.

Walsh, John P., and Nancy G. Maloney (2007). "Collaboration structure, communication media, and problems in scientific work teams." *Journal of Computer‐Mediated Communication* 12.2: 712-732.

Wennerås, Christine and Wold, Agnes (1997) 'Nepotism and sexism in peer review', Nature, 387: 341-43.

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi (2007). "The increasing dominance of teams in production of knowledge." *Science* 316.5827: 1036-1039.

Other References

Donovan, Claire (2011). Special Issue on the State of the Art in Assessing Research Impact. *Research Evaluation* 20 (3).

Holbrook, J. Britt (2009*). Social Epistemology: A Journal of Knowledge, Culture and Policy. Special Issue: US National Science Foundation's Broader Impacts Criterion*. 23 (3-4), July—December.

Laudel, Grit and Gloria Origgi (2006). Special issue on the assessment of interdisciplinary research. *Research Evaluation* 15: 1– 80.